PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 7 : C12N 15/86, 5/10, 7/04, 15/49, C07K 14/16, A61K 48/00, C12N 15/00, A01K 67/027

A2

(11) International Publication Number: WO 00/65076

(43) International Publication Date: 2 November 2000 (02.11.00)

(21) International Application Number: PCT/EP00/03765

(22) International Filing Date: 26 April 2000 (26.04.00)

(30) Priority Data:

99201306.0	26 April 1999 (26.04.99)	EP
00200171.7	18 January 2000 (18.01.00)	EP

(71) Applicant *(for all designated States except US)*: K.U. LEUVEN RESEARCH & DEVELOPMENT [BE/BE]; Groot-Begijnhof, Benedenstraat 59, B-3000 Leuven (BE).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: DEBYSER, Zeger [BE/BE]; Rega Institute for Medical Research, Minderbroedersstraat 10B, B-3000 Leuven (BE). DE CLERCQ, Erik [BE/BE]; Rega Institute for Medical Research, Minderbroedersstraat 10B, B-3000 Leuven (BE). CHEREPANOV, Peter [RU/BE]; Rega Institute for Medical Research, Minderbroedersstraat 10B, B-3000 Leuven (BE). PLUYMERS, Wim [BE/BE]; Rega Institute for Medical Research, Minderbroedersstraat 10B, B-3000 Leuven (BE).

(74) Agents: BIRD, William, E. et al.; Bird Goën & Co, Vilvoordsebaan 92, B-3020 Winksele (BE).

(81) Designated States: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

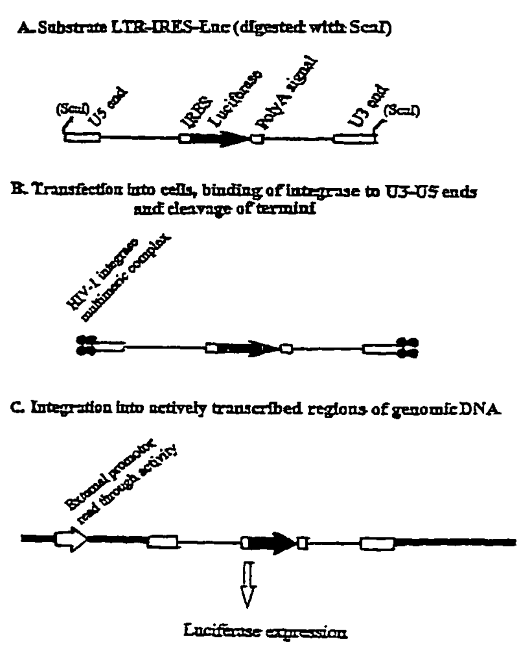
Published
Without international search report and to be republished upon receipt of that report.

(54) Title: SYNTHETIC GENE FOR EXPRESSING ACTIVE RETROVIRAL PROTEIN IN EUKARYOTES

(57) Abstract

The present invention features a synthetic gene or region of a gene which has an amended codon usage compared with the wild-type gene and which is for the high level expression of a retroviral protein in eukaryotic cells, the expressed retroviral protein having enzymatic activity in the eukaryotic cell. In addition, the invention features a synthetic gene or region of a gene encoding a retroviral enzyme or part of a retroviral enzyme normally expressed in a mammalian or other eukaryotic cell wherein at least one non-preferred codon in the wild-type gene encoding the enzyme has been replaced by a preferred codon encoding the same amino acid. The retroviral protein may be a protease, reverse transcriptase, integrase protein or a polyprotein gag-pol precursor thereof. In one embodiment the retroviral protein with enzymatic activity is a lentiviral protein. In other embodiments the enzymatically active protein is a *pol* enzyme. In more preferred embodiments, the enzymatically active protein is a lentiviral integrase. In an even more preferred embodiment the enzyme is an HIV enzyme. In more preferred embodiments the enzymatically active protein is HIV integrase. The present invention also includes a detection method for intracellular integrase using a promoterless reporter gene.

Principle of DIPR
Detection of integrase activity using a promoterless reporter gene



FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav	TM	Turkmenistan
BF	Burkina Faso	GR	Greece		Republic of Macedonia	TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's	NZ	New Zealand		
CM	Cameroon		Republic of Korea	PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

SYNTHETIC GENE FOR EXPRESSING ACTIVE RETROVIRAL PROTEIN IN EUKARYOTES

The present invention relates to the design of a synthetic gene for expressing retroviral proteins in eukaryotic cells especially mammalian cells as well as a synthetic gene, an expression vector containing the gene, eukaryotic cells stably harboring the gene, as well as methods of detection.

TECHNICAL BACKGROUND

Retroviruses are diploid positive strand RNA viruses that replicate through an integrated DNA intermediate. Typically, retroviruses comprise a protein-containing lipid envelope surrounding a protein-encapsulated core carrying the viral genome. Within the infected cell the retroviral genome is reverse-transcribed into double stranded DNA by a virally encoded reverse transcriptase enzyme that is part of the retroviral particle. The particle also includes other enzymes such as integrase. Integrase is the virus-encoded enzyme that is responsible for inserting the viral DNA copy into the chromosome of the host cell, a process referred to as retroviral integration. (For a review see Brown (1997), in Retroviruses, Cold Spring Harbor Laboratory Press USA, pp. 161-203). Integration is an essential step in the replication cycle of the human immunodeficiency virus type 1 (HIV-1), the causative agent of AIDS (LaFemina et al. (1992), J. Virol. 66: 7414-7419). Since no human counterpart is known to exist, integration has attracted a lot of attention as a potential new antiviral target. However, integrase inhibitor development has suffered from the lack of a relevant cellular integration assay; integrase activity is typically evaluated using artificial oligonucleotide-based test tube reactions. There is therefore a need to provide an intracellular integration assay.

Wild-type retroviral genomes contain at least three genes known as the *gag*, *pol* and *env* genes. The *gag* gene encodes internal core structural proteins, the *pol* gene encodes for certain enzymes such as protease, reverse transcriptase and integrase, and the *env* gene encodes the retroviral envelope glycoproteins. Integrases from different retroviruses vary in size from 30 to 46 kDa, are encoded by the 3'-end of the *pol* gene and are released from a *gag-pol* polyprotein precursor by proteolytic processing. The aminoterminal domain of integrase is characterized by a zinc finger (HHCC), is

universally conserved among all retroviruses, and is essential for in vivo integration. The central domain is the most conserved region with an essential DD35E motif involved in catalysis. This portion can catalyze the disintegration reaction in vitro. The carboxyterminal domain is referred to as DNA binding domain and shows the least sequence conservation. This fragment is required for 3'-end processing and integration. The active enzyme is thought to exist as a multimer wherein active domains can transcomplement inactive domains.

Transient expression of avian sarcoma-leukosis virus (ASLV) integrase in COS cells has been obtained previously (Morris-Vasios et al. (1988), J. Virol. 62: 349-353). A mouse cell line stably expressing the integrase of Rous sarcoma virus (RSV) has also been reported (Mumm et al. (1992), Virology 189: 500-510). Expression levels were not specified but appeared rather low. The integrase (IN) of HIV-1 has been expressed in *Escherichia coli* (*E. coli*) (Sherman and Fyfe (1990), Proc. Natl. Acad. Sci. USA 87, 5119-5123), insect cells using baculovirus (Bushman et al. (1991), Science 249: 1555-1558), and *Saccharomyces cerevisiae* (Caumont et al. (1996), Curr. Genet. 29: 503-510). In yeast integrase expression proved to be toxic in cells defective in DNA repair. High level expression of HIV-1 integrase in mammalian cells has remained elusive, in large part because expression of HIV-1 *gag* and *pol* proteins in general is Rev-dependent (Cullen (1992), Microb. Rev. 56: 375-395). In mammalian cells Rev-dependent expression of HIV-IN or HIV-IN fused to β-galactosidase or GFP has been reported previously (Faust et al. (1995), Biochem. Mol. Biol. Int. 36: 745-758; Kukolj et al. (1997), J. Virol. 71: 843-847; Pluymers et al., (1999), Virology 258: 327-332). However, expression levels, even after transient transfection, were always low. In the absence of Rev, multiple inhibitory or instability sequences (INS), also referred to as cis-acting repressor elements (CRS), in the mRNA interfere with protein expression. Potential mechanisms include: nuclear retention or mRNA instability. It was observed that mRNA containing CRS is trapped in the nuclei and that the inhibition of expression is at least partly due to the poor translocation of mRNA to the cytosol (Mikaelian et al. (1996), J. Mol. Biol. 257: 246-264; Borg et al. (1997), Virology 236: 95-103). Elements of the RNA processing machinery could be involved in nuclear trapping of mRNA that contains CRS. There is also evidence that several regions of the HIV-1 genome that contribute to the instability of the mRNA, have high AU contents. They may represent binding sites for cellular factors which contribute to mRNA

instability (Schneider et al. (1997), J. Virol. 71: 4892-4903). According to another hypothesis, mRNA containing inhibitory sequences fails to be translated efficiently without Rev. Whatever the mechanism of the observed inhibition, it is clear that inhibition occurs at the level of the mRNA and is due to some AU-rich regions. During the HIV replication cycle Rev interaction with the Rev responsive element (RRE) relieves the inhibition in a regulated manner (Schwartz et al. (1992), J. Virol. 66: 150-159). In this perspective, it is not surprising that by mutating some INS while preserving the coding function for *gag-pol* transcripts, efficient Rev-independent expression of viral particles has been obtained (Schneider et al. (1997), J. Virol. 71: 4892-4903). There is evidence that in the case of HIV gp120 mRNA poor translatability due to inefficient codon usage rather than mRNA instability is responsible for low level protein expression (Haas et al., 1996; Schneider et al. (1997), J. Virol. 71: 4892-4903).

US 5,811,270 (Grandgenett) describes a test tube method of analysis of concerted integration in which a viral integrase enzyme is first incubated with donor DNA molecules followed by incubation with target DNA molecules. The donor DNA has at least one unique restriction site for analysis of the concerted integration product. The described method is said to be useful for studying integrase such as screening of HIV-1 or HIV-2 integrase inhibitors as well as production of transgenic non-human animals and gene transfer. The integrase used is purified from virus particles and the activity is analyzed in the test tube, not intracellularly.

US 5,795,737, WO 96/09378, WO 97/11086 and WO 98/12207 all describe methods of producing a synthetic gene encoding a protein normally expressed in a mammalian cell whereby the synthetic gene is reported to overexpress the encoded proteins in mammalian cells. The known synthetic genes are constructed by replacing non-preferred codons or less preferred codons with preferred codons which encode the same amino acid by utilising the redundancy of the genetic code. Examples are given of synthetic *env* genes which encode envelope glycoproteins but there is no discussion of expressing a protein with enzymatic activity in the host cell. A method of designing a synthetic gene for the overexpression of a protein while maintaining its enzymatic activity is not derivable from the known teaching. There are a significant number of factors which may allow expression of a (retroviral) protein which fails to show intracellular enzymatic activity. The expressed enzyme may be defective for many

reasons of which intracellular inhibition of the enzyme and the need for the presence of another viral protein at the same time are but a few. Further, it is not obvious that an enzyme can be overexpressed, for example there may be some limiting factor such as poor solubility or cellular toxicity. On the one hand high level expression of a retroviral enzyme will be required to detect the enzymatic activity, on the other hand levels which are too high may cause protein precipitation or cellular toxicity. For any retroviral enzyme to be active in the cell an optimal intracellular concentration will be required. In case of failure the suggestion is to replace non-preferred codons to a certain percentage, e.g. 90%, 80%, 70% ..., but there is no precise teaching of how to select which codons are to be replaced. In particular, there is no indication that a specific nucleotide pair frequency is of relevance to high level gene expression. It is not conclusive that the mechanism of RRE-instability (in *env*) is the same as, or even related to the mRNA instability problem in *gag* and *pol*. In fact there is evidence that the mechanisms are different. Hence, it is not predictable that Rev-independent expression of an *env* gene may be extrapolated to cure the instability problem of *gag* and *pol* genes.

It is an object of the present invention to develop an efficient expression system for an enzymatically active retroviral protein, in particular HIV-1 integrase, in eukaryotic cells, especially mammalian cells.

It s a further object to provide a more efficient detection method for retroviral enzyme inhibitors.

It is a further object of the present invention to provide a design method for the construction of a gene encoding a retroviral protein with enzymatic activity.

A further object of the present invention is to provide an expression vector capable of delivering a gene to a target cell, in which cell the enzymatically active protein encoded by the gene is expressed.

SUMMARY OF THE INVENTION

The present invention features a synthetic gene or region of a gene which has an amended codon usage compared with the wild-type gene and which is for the high level expression of a retroviral protein in eukaryotic cells, the expressed retroviral protein having enzymatic activity in the eukaryotic cell. In addition, the invention features a synthetic gene or region of a gene encoding a retroviral enzyme or part of a retroviral

enzyme normally expressed in a mammalian or other eukaryotic cell wherein at least one non-preferred codon in the wild-type gene encoding the enzyme has been replaced by a preferred codon encoding the same amino acid. By "region of a gene with amended codon usage" is meant that it can be sufficient to change codons only in those parts of a gene that normally produce instability sequences (INS) or cis-acting repressor elements (CRS) in the transcribed mRNA of the gene.

By "retroviral protein or enzyme normally expressed in a mammalian or eukaryotic cell" is meant a protein or enzyme which is expressed in a mammalian or eukaryotic cell under disease conditions. These are genes which are encoded by a retrovirus (including a lentivirus) which are expressed in mammalian or eukaryotic cells post-infection.

In preferred embodiments, the synthetic gene is capable of expressing the retroviral enzyme at a level at least 200% of that expressed by the "natural" (or "native") gene in a mammalian or eukaryotic cell culture system.

The retroviral protein may be a protease, reverse transcriptase, integrase protein or a polyprotein gag-pol precursor thereof. In one embodiment the retroviral protein with enzymatic activity is a lentiviral protein. In other embodiments the enzymatically active protein is a *pol* enzyme. In more preferred embodiments, the enzymatically active protein is a lentiviral integrase. In an even more preferred embodiment the enzyme is an HIV enzyme. In more preferred embodiments the enzymatically active protein is HIV integrase. The enzymatic activity includes at least an integrase function, namely of promotion or stimulation of the integration of DNA fragments into host cell DNA, preferably the chromosome of the host cell. The integrase hereby is expressed on its own id est as a single component, independent of any retroviral components..

By "retroviral components" is meant the retroviral, specifically the lentiviral, and more specificially the HIV-1 regulatory and accessory proteins like Tat, Rev, Nef, Vpu, Vif, Vpr.

The invention also features a eukaryotic expression vector comprising the synthetic gene or region of a gene. The expression vector preferably includes a constitutive or an inducible or a tissue-specific promoter. Expression from the eukaryotic expression vector can be transient after transfection of the vector in a eukaryotic cell by any of suitable, e.g. established, transfection procedures. The vector may be any suitable vector such as a plasmid, a mammalian or insect virus. Expression

may also be permanent in a eukaryotic cell line stably harbouring the expression vector. The expression vector may be comprised in a packaging construct for producing retroviral particles for gene transfer. The retroviral particle may be a lentiviral particle.

Another aspect of the present invention features a eukaryotic cell line that harbours the synthetic gene or region of a gene. The cell line preferably expresses the retroviral enzymatically active protein using a constitutive, inducible or tissue specific promoter. The expressed retroviral protein shows enzymatic activity that can be measured for example by complementation of enzyme-defective viruses or in the case of an integrase by stimulation or the promotion of the insertion of DNA molecules into another DNA molecule, preferably the chromosome of the cell.

The present invention also includes a transgenic non-human animal harboring the synthetic gene or region of a gene. The expression of the gene or region of a gene may be induced at any moment using an inducible promoter or, alternatively, in desired tissues using a tissue-specific promoter.

The present invention also features a method for preparing a synthetic gene or region of a gene encoding an enzymatically active retroviral protein or part of such a protein. The method not only identifies and uses preferred codon usage but also, and moreover mainly, seeks to increase mRNA stability during expression. The method includes identifying a small group of genes from the total set of genes of a target eukaryotic cell which encode proteins which are naturally expressed easily and/or in high concentrations in the target cell. The small group may include 10 or less genes, more typically 5 or less genes. From the codon sequences of these identified genes, a preferred codon usage and a preferred nucleotide relationship or nucleotide pair frequency is identified. By preferred codon usage is meant that for a specific amino acid a specific codon is chosen as the preferred codon to encode the amino acid based on the high use of the preferred codon within the select group of genes. By a preferred codon relationship is meant the ratios of the various nucleotides and combinations of nucleotides to each other which commonly appear in genes of the target eukaryotic cell. One particular nucleotide relationship is the GC content or the GC nucleotide pair frequency. Using the preferred codon usage, non-preferred codons are identified in the natural gene encoding the enzyme and one or more of the non-preferred codons is/are replaced with a preferred codon encoding the same amino acid as the replaced codon. The replacement is biased to obtain the preferred nucleotide relationship or nucleotide

pair frequency, resulting in even better optimized conditions for expression in eukaryotes compared to the use of preferred codon usage only. The replacement may be made based on a random choice between alternative codons encoding the same amino acid at each position using a random number generator and biasing the choice of alternative codons based on the preferred codon usage to obtain the preferred nucleotide relationship or nucleotide pair frequency. In addition, the synthetic gene sequence may be edited by removing potential splice sites and to reduce the number of CpG methylation sites while keeping the overall nucleotide relationship or the nucleotide pair frequency close to the preferred one, e.g. keeping the GC content and codon usage close to the preferred one. GC content should be kept close to the preferred usage in the target cell, e.g. about 60% in mammalian cells. A preferred range for the GC content is 53 to 63%, more preferably 55 to 61% for expression of the gene in human cells. To provide efficient initiation of translation the Kozak consensus sequence (ANNATGG) may be added.

It is not necessary to replace all non-preferred codons with preferred codons. Increased expression may be accomplished even with partial non-preferred codon replacement with preferred codons. Under some circumstances it may be desirable to only partially replace non-preferred codons with preferred codons in order to obtain an intermediate level of expression.

By "synthetic gene" is meant a nucleotide sequence encoding a naturally occurring protein in which a portion of the naturally occurring codons has been replaced by other codons. For example, a non-preferred codon is replaced with a preferred codon encoding the same amino acid. However, by replacing codons to create a synthetic gene the expression in eukaryotic, e.g. mammalian cells (especially human cells) of a wide variety of genes (of eukaryotic, mammalian, prokaryotic or viral origin) can be increased compared to the expression of the naturally occurring gene. Thus, the invention includes improving the eukaryotic, especially a mammalian cell expression of a gene from any source by the codon replacement methods described herein.

By "vector" is meant a DNA molecule, derived, e.g., from a plasmid, or mammalian or insect virus, into which fragments of DNA may be inserted or cloned. A vector will contain one or more unique restriction sites and may be capable of autonomous replication in a defined host or vehicle organism such that the cloned sequence is reproducible. Thus, by "expression vector" is meant any autonomous

element capable of directing the synthesis of a protein. Such DNA expression vectors include mammalian plasmids and viruses.

By retroviral "packaging construct" or "packaging vector" is meant a plasmid-based or virus-based vector or construct, configured to encode for the proteins necessary for producing virus particles that are devoid of genomic RNA. In general, this implies providing the *gag*, *pol* and *env* gene products. Lentiviral packaging constructs of interest contain changes to the coding sequences of *gag* or *pol* proteins (i.e. synthetic genes) to enhance lentiviral protein expression and to enhance safety. For biosafety reasons, the packaging functions are often divided into two genomes, one which expresses the *gag* and *pol* gene and another expressing the *env* gene product. In packaging constructs in accordance with the present invention, regulators of gene expression such as the *Rev* gene product would no longer be required. Increased biosafety of these packaging constructs is based on a reduced risk for (homologous) recombination of these synthetic genes with their wild-type counterparts.

The invention also features synthetic portion of a gene which encodes a desired portion of the protein. Such synthetic gene fragments are similar to the synthetic genes of the invention except that they encode only a portion of the protein. The portion of the gene encodes a portion of the enzyme which has some enzymatic activity, e.g. it may have catalytic activity, for example, the synthetic gene may encode a catalytic core of an enzyme, e.g. it may be a part of reverse transcriptase.

The present invention also includes a detection method for intracellular integrase using a promoterless reporter gene. The reporter gene may be luciferase, GFP or an antibiotic selection marker (e.g. neomycin resistance). The reporter gene construct may be used as the substrate of the retroviral enzyme, e.g. integrase expressed from the synthetic gene be it in a stable cell line or in a transient mode after transfection of the expression vector, the retroviral enzyme, e.g. integrase being in accordance with the present invention.

The present invention may provide a synthetic gene and a method of designing and constructing the same to obtain efficient expression of a retroviral, in particular lentiviral enzyme such as integrase of the human immunodeficiency virus type 1 (HIV-1), or part of a retroviral enzyme in mammalian cells. The synthetic gene circumvents mRNA instability by increasing the GC content of the wild type integrase gene from 40% to 59%. The synthetic gene, cloned in a eukaryotic expression vector, provides

efficient expression of HIV-1 integrase in various mammalian cell lines. The amino terminus of the protein was as predicted by the sequence after removal of the first methionyl residue. Nuclear localization of the recombinant protein was evidenced by fluorescence microscopy. A 293T cell line stably expressing HIV-1 integrase was obtained. The functionality of integrase was proven by trans-complementation experiments. Lentiviral vector particles carrying the inactivating D64V mutation in the integrase gene, were obtained capable of stably transducing 293T cells when complemented in the producer cell line with integrase expressed from the synthetic gene. When the cell line that stably expresses integrase was infected with the defective virus particles, complementation of integrase function was observed. Transfection with a linear promoterless DNA substrate that contains a reporter gene behind an IRES and is flanked by HIV LTR ends, resulted in a reproducibly higher reporter signal in cells that express integrase. Since the increase in reporter gene activity was stable upon passaging of the transfected cells, it can be concluded that the integrase promotes insertion of the linear DNA substrate in the cellular chromosome. The fold increase of reporter signal with integrase expressed from a mutant synthetic gene, containing the D64V mutation, was considerably lower, indicating that the enzymatic activity of the enzyme was required. The established cellular integration system in accordance with the present invention facilitates the study of the interplay between host and viral factors during integration, the development of specific HIV integration inhibitors as well as the design of gene transfer systems.

The present invention, its advantages and embodiments will now be described with reference to the following figures and drawings.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1. Western blot analysis of transient expression of HIV-1 IN in 293T cells using different expression strategies. 293T cells were transiently transfected with the various expression vectors. At 48 hrs post transfection cell extracts were made using 1% SDS, 1 mM PMSF. Cell extracts representing 10 μg of total protein were separated by PAGE and blotted onto PVDF membranes. Detection was performed using polyclonal antibodies against HIV-1 integrase and the ECL+ detection system. Lane 1 contains 2.5 ng of recombinant and purified His-tagged HIV-1 integrase (HT-IN). The other lanes contain extracts after transfections with equal amounts of the

following plasmids: Lane 2, pCEP4; Lane 3, pCEP-IN; Lane 4, pCEP-IN-CTE; Lane 5, pCEP-IN-RRE + pEF-cREV; Lane 6 pCMV-IN^{s}.

Figure 2. Sequence and structure of the synthetic gene.

(A) Sequence of the synthetic DNA coding for pNL4-3 HIV-1 integrase. The amino acid sequence is shown in the single letter code. The restriction sites used in construction are boxed. The translation initiation site is underlined.

(B) A schematic representation of the structure of the synthetic gene. The following regions are indicated : the 5'- and 3'- untranslated regions (UTR) derived from β-globin mRNA, the Met-Gly dipeptide and the integrase open reading frame (ORF). The three domains of the integrase protein are shown: the Zinc finger motif (HHCC), the catalytic core and the DNA binding domain.

Figure 3. Western blot analysis of the 293T-derived cell line that stably expresses HIV-1 IN from the synthetic gene. 293T cells were transfected with pCMV-IN^{S} and a stable cell line was selected with HygromycinB. Cell extracts (10 µg of total protein) were separated by PAGE and blotted onto PVDF membrane. Detection was performed using polyclonal antibodies against HIV-1 integrase and the ECL+ detection system. Lane 1, 2.5 ng recombinant His-tagged HIV-1 integrase; Lane 2, extract of 293T cells; Lane 3, extract of 293T cells stably expressing IN (293T-IN^{S}).

Figure 4. Detection of integrase activity using a promoterless reporter construct (DIPR) Figs. 4A-C are schematic representations of the method of detection of integrase activity using a promoterless reporter gene.

DESCRIPTION OF THE ILLUSTRATED EMBODIMENTS

The present invention will mainly be described with reference to a synthetic gene for overexpressing HIV integrase in mammalian cells but the invention is not limited thereto but only by the claims.

It has long been known that expression of eukaryotic genes in prokaryotes can be optimised by designing synthetic genes with modified codon usage. Less established, although demonstrated, is the concept of increasing expression of eukaryotic genes in eukaryotic cells by modified codon usage. From bacteria it is known that few general rules apply (Makrides F. (1996), Microb. Rev 60: 512-538).

A retroviral enzyme such as integrase does not normally, during the infectious cycle, work as a soluble protein in the cytoplasm of a host cell.

Integrase is part of a large ill-defined nucleoprotein complex called the preintegration complex of which also reverse transcriptase, nucleocapsid, matrix protein, the viral DNA and other factors are part. It is not obvious that integrase on its own in the cytoplasm of a target cell is enzymatically active, for example, there may be cellular factors which inhibit activity or viral factors which are missing in this environment. Further, it is not obvious that integrase expressed as such will interact with artificial DNA substrates (see DIPR below). One aspect of the present invention is dissecting the preintegration complex to obtain a simple integrase-linear DNA interaction. One embodiment of the present invention is a method to detect and utilize the enzymatic activity of a retroviral, in particular a lentiviral enzyme, in particular integrase by itself in a eukaryotic cell.

Initially eukaryotic expression vectors encoding HIV-1 IN and IN-RRE were constructed based on the reasoning that co-expression of Rev in cells transfected with IN-RRE would increase expression levels of IN. However, in human cells transiently transfected with these expression vectors, little or no expression of IN was detected by either immunofluorescence microscopy or western blotting (Fig. 1). An alternative approach consisted of introducing the constitutive transport element (CTE) of simian retrovirus type 1 behind the integrase gene. Again, expression was barely detectable upon prolonged exposure of the blot and amounted to merely 40 ng per 10 x 10^6 transfected cells. The construction of a C-terminal fusion to green fluorescent protein (GFP) (GFP-IN) resulted in a more pronounced expression of wild-type HIV-1 integrase expressed in mammalian cells (Pluymers et al. (1999), Virology 258: 327-332). Rev co-expression was not required, in accord with Kukolj et al. (1997), J. Virol. 71: 843-847), who expressed integrase as a C-terminal fusion protein with β-galactosidase in the absence of Rev. The impact of the INS in the IN gene on protein expression levels was illustrated by a 5-fold decrease in expression levels of the GFP-IN construct compared to the parental GFP (Pluymers et al. (1999), Virology 258: 327-332). The present invention is based on a synthetic gene for HIV-1 integrase with an increased intrinsic mRNA stability. The use of such a synthetic gene resulted in high expression levels and concurrent enzymatic integrase activity as could be demonstrated via complementation tests.

In accordance with the present invention, an integrase gene was synthesised with an increased GC content resulting in high level expression of HIV-1 IN in various

mammalian cell lines. The enzyme was shown to complement defective integrase carried by HIV-1-derived vector particles and to act in trans on linear DNA substrates that are flanked by LTR fragments and encode a reporter gene.

DESIGN AND CONSTRUCTION OF THE SYNTHETIC INTEGRASE GENE

Synthetic genes have been constructed in the past to optimize expression of eukaryotic genes in bacteria based on the knowledge that codon usage in prokaryotes is quite different from that in eukaryotes. HIV (lentiviral) genes are not optimal for high level expression in eukaryotic cells. This is related to the mechanism HIV uses to circumvent the mRNA instability, namely Rev. During the replication cycle early mRNA transcripts will be spliced which results in expression of regulatory proteins such as Tat and Rev. Only late in the cycle, does Rev accumulation and Rev-RRE interaction block splicing and suppress AT-rich instability sequences resulting in unspliced transcripts encoding structural and enzymatic proteins. Whereas the synthetic gene in accordance with the present invention clearly augments protein expression in mammalian cells, which is a prerequisite to detect the functionality of the enzyme in the cell, in the context of replicating HIV the presence of a gene with an increased GC content may well interfere with the mechanism of regulation of gene expression and be detrimental for viral replication.

In accordance with an embodiment of the present invention a synthetic viral gene was designed for better optimized and more efficient expression in mammalian cells. The HIV-1 integrase gene has a GC content of 40% whereas highly expressed human genes on average have a GC content of 55-61%. Hence, the GC content is one aspect of the preferred nucleotide relationship or nucleotide pair frequency in accordance with the present invention. By employing the degenerative nature of the genetic code and selecting for the preferred codon usage in the synthetic gene, the GC code content of a synthetic gene encoding HIV integrase would be increased up to 66% without altering the amino acid sequence. However, this is not preferred in accordance with the present invention. First of all, in accordance with the present invention, the choice among the alternative codons was biased in favour of preferred triplets (codons) found in a small group of genes of the total human genome which express well/strongly, e.g. human β-globin, α-, γ-actin and EF2 genes (method of determining the preferred codon usage). In addition the bias was such as to approximate the

preferred nucleotide relationship or nucleotide pair frequency, i.e. within the range 53 to 63%, more preferably 55-61% for the GC content. In fact a GC content of 59% rather than 66% was achieved. The other rules for redesigning retroviral genes for eukaryotic expression are: (i) removal of potential splice sites, (ii) reduction of the number of CpG methylation sites, (iii) introduction of 5' and 3'-untranslated regions (UTR) of a mammalian mRNA (in our case from human β-globin), (iv) addition of an extra N-terminal peptide (Met-Gly for the examples given below) for efficient initiation of translation. As a result expression levels from the synthetic gene in various mammalian cell lines were at least 25-fold higher than from the natural integrase gene. Efficient expression was also obtained in yeast (*Pichia pastoris*) (data not shown).

In accordance with one embodiment of the present invention a gene is provided to achieve high level expression of HIV-1 integrase in human cell lines by maintaining the amino acid sequence of IN from the pNL4-3 clone of HIV-1 while adapting the nucleotide codon usage to the codon usage of constitutively and highly expressed human genes ("preferred codon usage"). A first version of an artificial IN reading frame was based on random choice between alternative codons at each position using a random number generator, biasing in favour of preferred triplets as found in the human β-globin, α-, γ-actin and EF2 genes. Next, the DNA sequence was substantially edited to remove potential splice sites and to reduce the number of CpG methylation sites, but keeping the overall GC content and codon usage close to optimal ("preferred nucleotide relationship" or "nucleotide pair frequency"). The final version of the synthetic gene (Fig. 2 or SEQ ID NO:1) contains fragments of the 5'- and 3'-untranslated regions from the β-globin mRNA. This gene encodes for wild type HIV-1 integrase with addition of the N-terminal Met-Gly dipeptide. The extra glycine codon completes the Kozak's consensus sequence (ANNATGG) required for efficient initiation of translation. In the synthetic gene the overall GC content is 59% compared to 40% in the wild type. The gene was constructed from six synthetic DNA fragments, each approximately 150 bp long, by stepwise cloning. It should be understood that various homologs of the gene shown in Fig. 2 or SEQ ID NO:1 are included within the scope of the present invention. Reapplication of the random number biasing procedure in accordance with the present invention would generate alternative sequences all of them coding for the same protein and all having a similar preferred nucleotide relationship or nucleotide pair frequency. All such synthetic gene homologs are included within the scope of the present

invention.

The synthetic gene includes modification to those described above, the following modifications and improvements of the synthetic gene are included within the scope of the present invention. For example, the leader peptide can be replaced affecting the efficiency of translation and potential myristoylation (e.g. for example, a Met-Ala variant has been constructed). The 5' and 3'-UTRs may be replaced by UTRs from other mammalian mRNAs to optimize the stability of the transcript. Mutations in the open reading frame are also included within the scope of the present invention whereby the canonical integrase sequences (e.g. HHCC and DD35E) are preferably left unchanged. A more soluble version can be made by introducing for example the F185K/ F185H mutations. Other mutations may induce increased or altered catalytic activity of the enzyme in the eukaryotic cell. For example, the present invention includes a variant synthetic gene with the D64V mutation, known to reduce drastically the enzymatic activity of integrase. Synthetic genes of integrase are included within the scope of the present invention in which the genetic information of domains of other proteins are added. These domains preferably add additional properties to the enzyme such as sequence specificity in DNA binding. Examples of methods of providing specificity to a gene encoding integrase are described in WO 96/37626, US 5,811,270 without describing the specific innovative aspects of the present invention.

The synthetic gene for HIV-1 integrase was designed to circumvent inhibition of gene expression induced by instability sequences (INS) in the wild type integrase gene. This approach can be applied to retroviral integrases in general. In particular the aforementioned design method may be used to redesign any retroviral viral gene encoding a protein with enzymatic activity for efficient expression in eukaryotes. In particular, the design method of synthetic genes in accordance with the present invention will boost eukaryotic expression for retroviral genes encoding a protein with enzymatic activity, especially lentiviral integrases and *pol* proteins in general. The particular approach of the present invention could also be applied to redesign *gag* genes, in which mRNA instability due to the presence of INS elements and not poor translatability, like for *env*, would be the problem. Although the role of Rev in suppressing the effect of INS is only well studied in the case of HIV-1, all other lentiviruses are known to encode proteins analogous to Rev. Likewise the human T-lymphotropic and bovine lymphotropic viruses (HTLVs and BLV) encode Rev. Simple

retroviruses such as Mason-Pfizer monkey virus and simian retrovirus-1 (SRV-1) contain a constitutive transport element (CTE) that promotes nuclear export of unspliced mRNA. It has been shown that CTE can functionally substitute for Rev interacting with RRE. In fact, a low level transient expression from a wild type integrase gene with a downstream CTE of SRV-1 has been obtained by us using the methods of the present invention. Since the design of a synthetic gene in accordance with the present invention abolishes any need for co-expression of Rev and presence of RRE or CTE in the construct, this approach can improve expression of retroviral enzymes in general and integrases in particular.

In creating mammalian expression vectors, various eukaryotic expression plasmids can be used. Expression can be under control of a constitutive promoter (for example hCMV and RSV) or an inducible promoter. Examples of (commercially available) inducible expression systems are the ecdysone-inducible and the tetracyclin-inducible (Tet-Off and Tet-On) expression systems. Tissue-specific promoters that limit expression in specific tissues may also be envisaged. Examples are the established neuron-specific promoters Thy-1 and enolase. Inducible promoters may limit cellular toxicity, although a cell line that stably expresses integrase was obtained. In transgenic non-human animals harbouring the synthetic gene, expression may be induced at a desired moment using an inducible promoter or in desired tissues using a tissue-specific promoter.

Transient and stable expression of HIV-1 integrase in 293T and HeLa cells

The synthetic gene for integrase (IN^s) was cloned into the expression vectors pCEP4 and pBK-RSV under control of the human cytomegalovirus (hCMV) and Rous sarcoma virus (RSV) promoters, respectively. Transient and stable expression of IN was obtained in both 293T and HeLa cell lines, as verified by immunoblotting (Fig. 1, 3) and indirect immunofluorescence (data not shown). In transfected 293T cells the expression levels from the hCMV promoter amounted to 10-20 μg of IN per 10 x 10^6 cells which is at least 25-fold higher than obtained with expression vectors that contain the unfused wild type HIV-1 integrase gene.

Transfection of 293T cells with the episomal expression vector pCEP-IN^S followed by selection with hygromycinB, resulted in a stable cell line, referred to as 293T-IN^S. Indirect immunofluorescence staining revealed that 80-90% of selected cells

produce integrase at detectable levels. The expression level, as estimated by quantitative immunoblotting, was about 0.5 µg of integrase per 10 x 10^6 cells. The reduced cell growth kinetics of 293T-IN^S (30-50% as compared to the parental 293T cell line) is suggestive of cellular toxicity of integrase in mammalian cells.

In HeLa cells integrase was found exclusively in the nuclei. In 293T cells transient transfections typically gave rise to an irregular, granular cytoplasmatic distribution of IN, probably due to precipitation of the protein. In the 293T cell line selected to stably express IN, nuclear localization of IN was evident. During the metaphase and anaphase steps of mitosis, IN remained stably associated with the chromosomes.

Solid phase N-terminal sequencing of integrase purified from transiently transfected 293T cells, revealed the following amino terminus: Gly-Phe-Leu-Asp-Gly-Ile-Asp-Lys. This is the sequence predicted by the synthetic gene, the starting methionine being removed post-translationally.

Functionality of IN^S

Complementation of IN-defective vector particles

To verify whether the integrase expressed from the synthetic gene in mammalian cells is enzymatically active, the ability of IN to complement integrase-defective HIV-derived lentiviral vectors was tested. HIV-1-derived lentiviral vectors have been developed by Naldini et al. and Zufferey et al. (Naldini et al. (1996), Science 272: 263-267; Zufferey et al. (1997), Nature Biotechnol. 15: 871-875). Pseudotyped lentiviral vector particles are produced by transfecting 293T cells with a packaging plasmid encoding viral *gag* and *pol* proteins, a plasmid encoding the envelope of vesicular stomatitis virus and a plasmid encoding a reporter gene flanked by two long terminal repeats (LTRs). The first generation packaging plasmid pCMVΔR8.2, containing all HIV genes except for *env* and the transfer vector pHR'-CMVLacZ were used to produce wild type vector (WT vector). Integrase-defective virus particles (D64V vector) were produced using pCMVΔR8.2IN(D64V) (Naldini et al. (1996), Science 272: 263-267). The D64V mutation in the integrase gene is known to abolish integrase activity, without affecting any other step of the infection (Leavitt et al. (1993), J. Biol. Chem. 268: 2113-2119; Leavitt et al. (1996), J. Virol. 70: 721-728). The transducing titer of the D64V vector in 293T cells was 20-fold lower than the titer

of WT vector (Table 1). This is in good agreement with previously reported results (Naldini et al. (1996), Science 272: 263-267). The observed "background" expression after D64V transduction, is mostly due to transcription from non-integrated circularized viral DNA since β-galactosidase expression after D64V transduction is reduced drastically upon passaging the cells (Table 1). Nevertheless, in some of the transduction experiments 1 or 2 galactosidase-positive colonies were observed. A residual transducing activity of D64V virus was observed before (Gaur and Leavitt (1998), J. Virol. 72: 4678-4685). It is possible that this integration is independent of the viral integrase.

Complemented vectors (C IN) were produced after quadruple transient transfection of producer cells, including pCEP-IN^S, the expression vector containing the synthetic gene. The transducing activity was restored up to 30% with C IN (Table 1). Complementation was due to stable integration, since an equal proportion of galactosidase-positive colonies was counted after multiple passages of the transduced cells. The principle of trans-complementation of IN-defective virus was shown previously, using VPR-IN fusion expression constructs (Fletcher et al. (1997), EMBO J. 16: 5123-5138). The transducing activity of catalytic domain mutants of IN was restored up to 20% by transcomplementation with VPR-IN. However, in the absence of VPR, the expression construct for wild type integrase, only achieved 0.04% complementation efficiency (Fletcher et al. (1997), EMBO J. 16: 5123-5138). The synthetic gene in accordance with the present invention, in the absence of VPR, results in a complementation activity that is 750-fold more pronounced.

Moreover, evidence for trans-complementing activity of integrase expressed from the synthetic gene in target cells was also obtained (Table 1). Transduction of IN-expressing 293T cells with IN-defective virus particles, resulted in a higher transduction efficiency as compared with the parental 293T cells. After passaging the transduced cells, the difference became even more pronounced. This points to a catalytic interaction of integrase present in the receptor cell with the pre-integration complex of the incoming vector. For the wild type and the complemented vectors increased transduction efficiencies were obtained as well. This may suggest that the amount of active integrase present in the viral particle is dose-limiting or that integrase present in the target cell neutralizes inhibitory host factors.

Detection of integrase activity using a promoterless reporter gene (DIPR).

Integration of HIV in the chromosome does not show strict sequence-specificity, although a weak consensus was found for the integration sites (Carteau et al.(1998), J. Virol. 72, 4005-4014). It is commonly accepted although not formally proven, that retroviral integration is favored in open chromatin near or within active transcription units (Rohdewohld et al. (1987), J. Virol. 61: 336-343; Scherdin et al. (1990), J. Virol. 64, 907-912; Vijaya et al. (1986), J. Virol. 60: 683-692; Carteau et al. (1998), J. Virol. 72, 4005-4014). The design of a promoterless reporter substrate for measuring integrase activity in cell culture, is based on this finding (Figs. 4A - C). In accordance with an embodiment of the present invention a method is proposed in which read-through transcription of the integrated promoterless reporter gene will occur when inserted within an actively transcribed region of the chromosome. The construct designed is a linear DNA fragment, flanked by the 200 bp terminal fragments of the HIV LTRs that provide the integrase recognition sites. The marker gene may encode luciferase, for instance. The presence of an IRES (internal ribosome entry site) in front of the open reading frame of luciferase, directs cap-independent translation of mRNA transcripts (Fig. 4A).

After transfection with this DIPR substrate (Fig. 4B, C), luciferase activity measured in 293T-IN^S cells was always 4 to 10 times higher than in the parental 293T cells (Table 2). In the DIPR assay activity of the D64V mutant integrase was reduced compared to the wild type integrase (data not shown). These results point to an activity of the intracellularly expressed integrase (expressed by the synthetic gene in accordance with the present invention) (Fig. 4C). Sequencing of integrated linear DNA molecules in 293T cells transiently expressing integrase from the synthetic gene using Alu-PCR, revealed the characteristic removal of the 3' GT dinucleotide in 10% of integrants. In control cells not expressing integrase none of the DNA insertions showed this hallmark.

Applications

An embodiment of the present invention includes the construction of an efficient eukaryotic expression vector for a retroviral enzyme, e.g. HIV-1 integrase, based on the creation of a synthetic gene. Expression from the eukaryotic expression vector can be transient after transfection of the plasmid in a eukaryotic cell by any of

established transfection procedures. Expression may also be permanent in a cell line stably harbouring the expression vector. An important aspect of the present invention and its applications is the functionality of an expressed retroviral enzymatically active protein, as opposed to mere the high level expression of an enzymatically inactive retroviral protein.

Intracellular integrase test for the evaluation of integrase inhibitors

An embodiment of the present invention includes assays for evaluating integrase activity in cells transfected with a DNA substrate that is flanked by fragments of HIV LTR, a so-called mini-HIV. In both assays data point to enzymatic activity of IN.

In *DIAS* (detection of integrase activity through antibiotic selection) test, a resistance gene to a cytotoxic drug is present in the mini-HIV DNA. The presence of IN in the transfected cell augments stable insertion of the resistance gene in the chromosome. Scoring is performed by comparing the residual number of colonies resistant to the cytotoxic agent in comparison with cells transfected with heterologous DNA.

In *DIPR* (detection of integrase activity using a promoterless reporter gene), a reporter gene (luciferase) without promoter is present downstream of an internal ribosome entry site (IRES) in mini-HIV (Fig. 4A). The presence of IN in the transfected cell (Fig. 4B) augments stable insertion of the reporter construct in the host chromosome in close proximity to a cellular promoter (Fig. 4C). Scoring is performed by measuring enzyme activity expressed from the promoterless marker gene, e.g. luciferase. The latter assay is highly amenable to evaluation of integrase inhibitors in cell culture in a microtiter plate format, adaptable for high throughput screening. Potential integrase inhibitors would result in the absence of or a significant decrease in the level of detectable signal from the promoterless marker gene.

Such an assay in accordance with the present invention involves screening test inhibitory compounds from large libraries of synthetic or natural compounds. Synthetic compound libraries are commercially available from, for example, Maybridge Chemical Co. (Trevillet, Cornwall, UK), Comgenex (Princeton, NJ), Brandon Associates (Merrimack, NH) and Microsource (New Milford, CT). A rare chemical library is available from Aldrich Chemical Company, Inc. (Milwaukee, WI).

Alternatively, libraries of natural compounds in the form of bacterial, fungal, plant and animal extracts are available from, for example, New Chemical Entities, Pan Laboratories, Bothell, WA or MycoSearch (NC), Chiron, or are readily producible. Plant extracts may also be obtained form the University of Ghent, Belgium. Additionally, natural and synthetically produced libraries and compounds are readily modified through conventional chemical, physical, and biochemical means.

Tool for non-viral cellular gene delivery.

Cell lines that express integrase from a synthetic gene do have greater propensity to integrate foreign DNA, flanked by LTR fragments. These cell lines are thus more transducible. An embodiment of this invention is the creation of eukaryotic cell lines (or cell culture systems) that are highly transducible (at least 200% compared to the parent cell). Embodiments of the present invention also include applications in transgene technology to increase the efficiency of (non)homologous recombination in ES cells be it by transient expression from a plasmid or after induced expression of the retroviral integrase in ES cells transgenic for the synthetic gene. The synthetic gene in accordance with the present invention may be brought into cells by any transfection agent or method (e.g. electroporation or lipofection) and may result in the stable integration of DNA in the chromosome.

Retroviral (lentiviral) vector packaging construct

From the complementation experiment it is clear that integrase expressed from the synthetic gene in the producer cell can complement integrase-defective lentiviral virus particles encoded by a packaging plasmid and thus can substitute for the protein expressed by the packaging construct. It follows that in an expression vector based on one or more synthetic gene(s) for a lentiviral *gag-pol* gene, the synthetic gene(s) can substitute for the natural gene(s) in the packaging constructs resulting in Rev-independent high level protein expression. The present invention includes a packaging construct based on non-lentiviral complex retroviruses in which protein expression is dependent on a Rev homologue such as Rex in the case of HTLV-I. Lentiviral vectors per se, capable of transducing a non-dividing cell, are known in the art (see Naldini et al. (1996), Science 272: 263-267, Zufferey et al. (1997), Nature Biotechnol. 15: 871-875). Generally the vectors are plasmid-based or virus-based, and are configured to

carry the essential sequences for incorporating nucleic acid, for selection and for transfer of the nucleic acid in the host cell. *Gag*, *pol* and *env* genes of interest are known in the art. Briefly, a first vector can provide a nucleic acid encoding a viral *gag* and a viral *pol*, and a second vector can provide a nucleic acid encoding a viral *env* gene product to produce a packaging cell. Packaging cells or cell lines supply in trans the proteins necessary for producing infectious virions, themselves being incapable of packaging endogenous viral genomic nucleic acids (Watanabe & Temin (1983), Molec. Cell Biol. 3(12): 2241-2249; Mann et al. (1983), Cell 33:153-159; Embretson & Temin (1987), J. Virol. 61(9): 2675-2683). Introducing a vector providing a heterologous gene, the transfer vector, into such packaging cells yields producer cells which release infectious particles carrying the foreign gene of interest. Methods for transfection or infection are well known by those skilled in the art. After cotransfection of the packaging vectors and the transfer vector to the packaging cell or cell line, the recombinant vector is recovered from the culture media and titered by standard methods used by those of skill in the art.

The foreign or heterologous gene carried by the transfer vector can be any nucleic acid of interest which can be transcribed, but preferably is a nucleic acid encoding for a polypeptide of therapeutic benefit or of interest for gene therapy. The *env* gene in the (second) packaging vector can be derived from any virus, including retroviruses, and is preferably amphotropic, allowing transduction of cells of human and other species, and is preferably under control of non-endogenous regulatory sequences. Vectors can be made target-specific through linkage of the *env* protein with an antibody or a ligand for a particular receptor of a particular cell-type (cell-targeting).

Design of the gag-pol synthetic gene is based on a method to circumvent mRNA instability associated with these wild-type genes. Preferentially, the method used by us to create an expression construct for high level and Rev independent eukaryotic expression of active HIV-1 integrase is employed. Further construction of the vectors of the present invention, whereby natural *gag-pol* genes are replaced respectively by the synthetic genes of the present invention, employ standard ligation and restriction techniques which are well understood in the art (see Maniatis et al, in Molecular Cloning: a Laboratory Manual, Cold Spring Harbor Laboratory, N.Y., 1982).

Biosafety requires measurements to reduce the risk of generating recombinant replication competent retroviruses (RCR) as much as possible. Dividing the packaging

functions into two genomes, one which express the *gag* and *pol* gene and another expressing the *env* gene product help to minimize the likelihood of generating RCR. That approach minimizes the ability for co-packaging and subsequent transfer of the two-genomes, as well as significantly decreases the frequency of recombination due to the presence of three retroviral genomes in the packaging cell to produce RCR. To render any possible recombinations non-functional, mutations (Danos & Mulligan (1988), Proc Natl. Acad. Sci 85: 6460-6464) or deletions (Bosselman et al. (1987), Molec. Cell Biol. 7(5):1797-1806; Markowitz et al. (1988), 62(4):1120-1124) can be configured within the undesired gene products. Deletion of the 3'LTR of both packaging constructs will further reduce the likelihood to form functional recombinants. US5,994,136 describes the production of lentiviral vectors with an even more remote possibility of generating replication competent lentiviruses by functionally deleting the *tat* gene, which is encoding for a regulating protein that promotes viral expression through a transcriptional mechanism. The likelihood of recombination between the transfer vector, that still contains natural genetic information of the lentivirus like part of the *gag* gene, and synthetic packaging genes will be considerably reduced, further improving the biosafety of the lentiviral vectors. DNA sequence mismatching, as induced by the replacement of nucleotides in the third position compared to the natural gene, seems to present a considerable barrier to homologous recombination in a wide variety of species. It is therefore also unlikely that contaminating or endogenous HIV virus particles would exchange the natural integrase gene for the synthetic one through recombination.

Experimental Procedures

DNA constructs

Construction of integrase expression plasmids

The open reading frame of IN from the HIV-1 clone HXB2 was PCR amplified using Pfu DNA polymerase (Stratagene, Cambridge, UK) with the primers 5'-CCCCCAAGCTTGCCAGCCATGTTTTTAGATGGAATAGATAAGG and 5'-CCCGCTC*GAG*CTTTCCTTGAAATATACATATGGTG and subcloned in pCEP4 (Invitrogen, Leek, The Netherlands), resulting in pCEP-IN. The absence of mutations was verified by DNA sequencing. The RRE sequence of HIV-1, clone HXB2, was PCR

amplified using the primers 5'-TTCCGCTC*GAG*TAGCACCCACCAAGGCAAA*GAG* and 5'-TCGCGGATCCAAGGCACAGCAGTGGTGCAAATG. The PCR fragment was subcloned in the sense orientation downstream of the integrase gene in pCEP-IN to produce pCEP-IN-RRE. The CTE sequence (obtained from plasmid pS12; Taberno et al., 1996, J. Virol. 70: 5998-6011) was cloned in pCEP4 in the correct orientation, followed by the insertion of the integrase gene upstream of the CTE. This resulted in the plasmid pCEP-IN-CTE. The construction of pGFP-IN is explained in Pluymers et al. (1999), (Virology 258: 327-332). The Rev expression plasmid, pEF321-cREV, was provided by Sandoz Forschungs Institut, Vienna, Austria. PCR amplification and plasmid construction employed standard techniques like standard ligation and restriction techniques and conditions which are well understood in the art (see Maniatis et al, in Molecular Cloning: a Laboratory Manual, Cold Spring Harbor Laboratroy, N.Y., 1982).

Assembly of the synthetic gene

The restriction sites NheI, PstI, BamHI, NaeI, NarI (indicated in Fig.2) divide the sequence of the synthetic gene into 6 fragments each approximately 150 bp long that correspond to the sequences 1-149, 144-306, 301-456, 451-623, 618-776, 771-930 (Fig. 2). Each of the fragments was constructed separately by annealing and extending two partially complementary oligonucleotides (85-95 nt long, PAGE-purified and 5'-phosphorylated, synthesized by Gibco BRL Life Technologies, Merelbeke, Belgium) using Sequenase (Amersham-Pharmacia, Buckinghamshire, UK) . Each fragment was cloned into the EcoRV site of the vector pBluescript KS(+) (Stratagene, La Jolla, CA). The sequence errors found in the resulting clones were repaired using either the Stratagene Quick Change procedure with *Pyrococcus furiosus* (Pfu) polymerase (for a base substitution in the fragment 451-623) or PCR (for deletions in the terminal regions of the fragment 1-149). The full 930 bp sequence was built by stepwise assembly of the fragments similar to the method described in WO98/12207. Choice of the cloning vector (pBluescript KS or SK) at each step was dictated by toxicity of the IN coding DNA. Finally, the two halves of the IN gene (1-451 and 452-930) were ligated together and cloned into pBluescript KS(+) resulting in pIN^{S}.

Construction of mammalian expression vectors for IN^{S}

The plasmid pIN^S was digested by EcoRI and treated with T4 DNA polymerase followed by restriction with XhoI. The 1 kb fragment carrying the IN^S gene was cloned between the PvuII and XhoI sites of pCEP4 (Invitrogen, Leek, The Netherlands) resulting in pCMV-In^S. pCEP4 is an episomal mammalian expression vector containing the human cytomegalovirus (hCMV) immediate early enhancer/promoter. The Epstein Barr virus replication origin (oriP) and nuclear antigen (encoded by the EBNA-1 gene) permit extrachromosomal replication in human, primate and canine cells. A hygromycin resistance gene is present, permitting selection of stably transduced clones by hygromycinB (GIBCO BRL). The same 1 kb fragment was also cloned between NheI and XhoI sites of the pBK-RSV expression vector (Stratagene) (the NheI cohesive end of the vector DNA was filled in using T4 DNA polymerase) resulting in pRSV-IN^S. In this vector expression of IN^S gene is driven by the promoter of Rous sarcoma virus (RSV). The presence of the neomycin resistance gene allows selection of stably transduced clones by geneticin (G418) (GIBCO BRL).

Construction of the substrate for the DIPR assay

The DNA substrate for the DIPR assay was obtained by linearization of pLTR-IRES-Luc with ScaI. This plasmid was constructed in the following way. First, the 350 bp KpnI/EcoRI fragment of pU3U5 (Cherepanov et al.(1999), Nucleic Acids Research 27: 2202-2210) containing the terminal U3 and U5 regions of the HXB2 HIV-1 LTRs was cloned between the KpnI and EcoRI sites of pUC19 resulting in pUC-LTR. Then the ScaI site occurring in the ampicilline resistance gene of pUC19 was destroyed by partial digestion of pUC-LTR with ScaI and insertion of a fragment containing the kanamicine resistance gene from the Tn5 transposon yielding pUC-LTR-kan. Finally, 7.5 kb pLTR-IRES-Luc was obtained by cloning the BamHI/PstI fragment of pBIR (Martinez-Salas et al. (1993), J. Virol. 67, 3748-3755) carrying the IRES-luciferase gene cassette, made blunt with T4 DNA polymerase (Gibco BRL), into the SmaI site of pUC-LTR-kan.

Cell culture

HeLa and 293 cells were obtained from American Type Culture Collection. HeLa and 293 cells were grown in Dulbecco's modified Eagle's medium (DMEM) (GibcoBRL) supplemented with 10 % FCS, 0.12 % (v/w) sodium bicarbonate

(GibcoBRL), 2 mM glutamine (GibcoBRL) and 20 µg/ml gentamycin (GibcoBRL) at 37°C in 5 % CO_2 humidified atmosphere. 293T cells (obtained from Dr. O. Danos, Evry, France) express SV40 large T antigen and were grown in DMEM (GibcoBRL) with glutamax supplemented with 10% fetal calf serum, 45 U/ml penicillin G (Serva, Heidelberg, Germany) and 45 µg/ml streptomycin sulphate (Sigma-Aldrich, Bornem, Belgium).

293 and 293T cells were transfected using polyethylenimine (PEI) (Abdallah et al. (1997), Hum. Gene Therapy 7:1947-1954). Polyethylenimine Mw ~ 25.000 was from Sigma-Aldrich (Bornem, Belgium). Cells were grown to 50-70 % confluency in DMEM with glucose, glutamax and 10 % fetal calf serum (FCS) (Gibco BRL). Medium was replaced by medium containing 1 % FCS 3 hours before transfection. Mixture of DNA and PEI was added to cells in a minimal volume of medium. Next day the medium was changed to DMEM containing 25 mM HEPES. Transformation efficiency obtained in this way was 50-80 %. HeLa cells were routinely transfected by electroporation. The cells were first trypsinized at 80 % confluency and pelleted by low speed centrifugation. The cells were then resuspended at a density of 2 x 10^6 cells/ml in growth medium; 0.5 ml of this solution was aliquoted into 4 mm cuvettes (Eurogentec, Seraing, Belgium) and 20 µg DNA was added to the cell suspension. After the electric pulse (10 µF, 250 V), cells were allowed to rest for 10 min at room temperature before dilution into growth medium..

To establish stable cell lines expressing the IN^S gene, cells transfected with pRSV-IN^S or pCMV-IN^S were cultured in the presence of 500 µg/ml geneticin (G418) or 200 µg/ml hygromicin B (both from GIBCO BRL), respectively. Expression of IN was assessed by western blotting and/or indirect immunofluorescence.

Western blotting and immunofluorescence

For western blotting and indirect immunofluorescence rabbit polyclonal antibodies directed against recombinant His-tagged HIV-1 IN were produced in house and were purified using a 1 HiTrap rProteinA column (Pharmacia Biotech, Uppsala, Sweden) according to established procedures (Ausubel et al. (1987), Current protocols in molecular biology, John Wiley & Sons, New York). Western blotting was performed using PVDF membranes (Bio-Rad), the ECL+ chemiluminescent detection system (Amersham-Pharmacia) and HRP-conjugated goat anti-rabbit antibodies (Bio-rad).

Dilutions used were 1:30000 for the primary antibody and 1:20000 for the secondary antibody. Detection limit was 0.1-0.5 ng of recombinant integrase. Total protein concentration was determined on cells lysed with 1%SDS/1 mM PMSF (Sigma), using the BCA protein assay (Pierce, Illinois USA). For western blot analysis 10 μg of total protein was evaluated.

For detection of IN expression *in situ* by indirect immunofluorescence microscopy, cells were grown on glass slides (HeLa cells) or in permanox chamber slides (GIBCO BRL) (293T cells). After 24-48 hrs, cells were washed with phosphate buffered saline (PBS) supplemented with 1 mM Mg^{2+} and 0.5 mM Ca^{2+}(PBS+), fixed in 100% methanol and blocked with 10% foetal calf serum (FCS) in PBS+. Incubations with antibodies were carried out at 37°C in blocking solution. The primary antibody (rabbit anti-IN) was diluted 1:20 to 1:80; the secondary FITC-conjugated swine anti-rabbit antibody from Dako (Glostrup, Denmark) was diluted 1:40. Nuclear staining was performed with 1 μg/ml 4', 6-diamidino-2-phenylindole (DAPI) (Sigma) in methanol. Fluorescence microscopy was performed with a Leitz microscope (Wetzlar, Germany) using filter blocks I2 (FITC) or A (DAPI).

Detection of integrase activity using a promoterless reporter gene (DIPR)

293T and 293T-IN^S cells were seeded in six-well plates at a density of 10^6 cells/well 24 hr before transfection. Five μg of DNA was transfected per well using PEI. 48 hr post-transfection, 5 x 10^5 cells were lysed to determine the luciferase activity using the Luciferase Assay System™ (Promega Benelux, Leiden, The Netherlands) and the Lumicount™ (Packard, Meriden, CT). The protein concentration of the lysate was determined using the Bradford method (Bio-Rad protein assay, Bio-Rad, Hercules, CA). The relative luciferase activity was calculated by dividing the luminescence values by the protein concentration.

Lentiviral vectors

Lentiviral vector production

HIV-1-derived vector particles, pseudotyped with the envelope of vesicular stomatitis virus (VSV), were produced by transfecting 293T cells with a packaging plasmid encoding viral *gag* and *pol* proteins (pCMVΔR8.2), a plasmid encoding the envelope of vesicular stomatitis virus (pMDG) and a plasmid encoding a reporter gene

flanked by two long terminal repeats (LTRs) (pHR'-CMVLacZ). The first generation packaging plasmid, containing all HIV genes except for *env*, and the transfer vector were a kind gift from Dr. O. Danos (Généthon, France). For transfection of a 10 cm dish of 293T cells, a 700 µl mixture of three plasmids was made in 150 mM NaCl: 20 µg of vector plasmid, 10 µg of packaging construct and 5 µg of envelop plasmid. To this DNA solution 700 µl of a PEI solution (110 µl of a 10 mM stock solution in 150 mM NaCl) was added slowly. After 15 min at room temperature, the DNA-PEI complex was added dropwise to the 293T cells in DMEM medium with 1% FCS. After overnight incubation, medium was replaced with medium containing 10% FCS. Supernatants were collected from day two to five post-transfection. The vector particles were sedimented by ultracentrifugation in a swinging-bucket rotor (SW27 Beckman, Palo alto, CA) at 25, 000 rpm for 2 hr at 4°C. Pellets were redissolved in PBS resulting in a 100-fold concentration. Different viral stocks were normalized based on p24 antigen content (HIV-1 p24 Core Profile ELISA, DuPont, Dreieich, Germany) for use in complementation assays.

Complementation experiments

Integrase-defective virus particles were produced using pCMVΔR8.2IN(D64V), obtained from Dr. D. Trono, (Geneva, Switzerland) as packaging plasmid (Naldini et al. (1996), Science 272: 263-267). Complemented vectors were produced by expressing integrase from pCEP-IN^S in 293T cells after quadruple transient transfection. Vector preparations were normalized for p24 antigen count Vector was added to target cells in the presence of 2 µg/ml polybrene and left overnight. After removal of vector, cells were incubated for an additional 36 hrs. Cells were washed with PBS, fixed with 0.75% formaldehyde/0.05% glutaraldehyde in PBS, and stained with freshly prepared X-gal substrate (5 mM potassium ferrocyanide, 5 mM potassium ferricyanide, 2 mM $MgCl_2$ and 100 µg/ml 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (x-gal) (Biotech Trade & Service Gmbh, St. Leon-Rot, Germany) in PBS) at 37°C overnight. Each transduction experiment was done in duplicate in a 96-well plate. Transduction efficiency was determined by counting the number of blue cells 48 hrs after infection in one of the wells, whereas the cells in the duplicate well were splitted 1:2. Half of the sample remained in the well and was stained at confluency (passage 1) whereas the other half was cultured in a 48-well plate. At confluency, these cells were again splitted

1:2. Finally, cells were brought in a 24-well plate and grown to confluency (passage 3, dilution 1:8). After staining, the efficiency of stable transduction was measured by counting blue colonies.

Tables

Table 1. Complementation of integrase-defective lentiviral vector particles

Relative transduction efficiency[1]

Cells	*Passage*	*WT vector*	*D64V vector*	*C IN*
293T				
	# 0	1.00	0.048	0.303
	# 3	1.00	0.007	0.320
293T IN^{S}				
	# 0	1.565	0.09	0.510
	# 3	1.88	0.045	0.75

[1]Transduction efficiency is determined by counting galactosidase-positive cells (# 0) or colonies of galactosidase-positive cells (# 3) relative to transduction efficiency obtained by WT vector in 293T cells. Results of transduction by WT vector, D64V IN-defective vector and D64V vectors complemented with IN in the producer cells, are shown. Cells were infected with normalized amounts of vector. Transduction was done both in 293T cells and in 293T cells that are stably expressing IN. Average numbers for two separate experiments are shown.

Table 2. Detection of integrase activity using a promoterless reporter gene (DIPR)

		Luciferase activity (Relative units)		
Experiment	Cell line	Blank[a]	LTR-IRES-Luc[b]	LTR-IRES-Luc + pCMV-IN^{S} [c]
A	293T	1	47 ± 1	-
	293T-IN^{S}	1	487 ± 119	-
B	293T	1	130 ± 24	489 ± 169
	293T-IN^{S}	1	499 ± 38	990 ± 183

[a]Relative background luciferase activity in cell lines

[b] 293T and 293T-IN^{S} were transfected with equal amounts of linearized pLTR-IRES-Luc. under experimental conditions A. In experiment B total DNA concentration was equalized with parental vector pCEP4.

[c]293T and 293T-IN^{S} were transfected with linearized pLTR-IRES-Luc and 2 μg of pCMV-IN^{S}.

CLAIMS

1. A detection method for intracellular integrase activity using a promoterless reporter gene.

2. The detection method according to claim 1, wherein the reporter gene may be luciferase, GFP or an antibiotic selection marker.

3. The detection method according to claim 1 or 2, wherein a reporter gene construct is generated from the reporter gene and the construct is used as the substrate of the enzymatically active retroviral protein expressed from the synthetic gene in accordance with claims 17 to 26.

4. Packaging construct for a lentiviral or complex retroviral vector based on a synthetic *gag* or *pol* gene.

5. Packaging construct according to claim 4, wherein the synthetic gene is the synthetic gene in accordance with any of claims 17 to 26.

6. A method of transfecting a eukaryotic cell using the expression vector in accordance with any of claims 27 to 29.

7. A eukaryotic cell line harboring the synthetic gene or region of a gene in accordance with any of the claims 17 to 26.

8. The eukaryotic cell line according to claim 7, wherein the retroviral enzymatically active protein is expressed using a constitutive, inducable or tissue specific promoter.

9. The eukaryotic cell line according to claim 7 or 8, wherein the expression is stable.

10. A transgenic, non-human, animal harboring the synthetic gene or region of a gene in accordance with claims 17 to 26.

11. The transgenic animal according to claim 10, wherein the expression of the synthetic gene or region of a gene is induced by an inducable promoter or by a tissue-specific promoter.

12. The transgenic animal according to claim 10 or 11, wherein the animal is a mammal.

13. A method for preparing a synthetic gene or region of a gene encoding a retroviral protein or part of such a protein which is enzymatically active in a target eukaryotic cell, comprising the steps of:
1) identifying a group of genes from the total set of genes of the target eukaryotic cell which encode proteins which are naturally expressed easily and/or in high concentrations in the target cell;
2) determining the codon sequences of these identified genes and from these sequences a preferred codon usage and a preferred nucleotide pair frequency;
3) using the preferred codon usage, identify the non-preferred codons in the natural gene encoding the enzymatically active protein;
4) replacing one or more of the non-preferred codons with one or more preferred codons encoding the same amino acids as the replaced codons while biasing the replacement to obtain the preferred nucleotide pair frequency.

14. The method according to claim 13, wherein the replacement step is carried out based on a random choice between alternative codons encoding the same amino acid at each position using a random number generator and biasing the choice of alternative codons based on the preferred codon usage to obtain the preferred nucleotide pair frequency.

15. A method for gene transfer in a eukaryotic cell expressing the synthetic gene or region of the gene in accordance with any of the claims 17 to 26.

16. A method according to claim 15, wherein the synthetic gene is transiently expressed or is stably integrated in said cell.

17. A synthetic retroviral *gag* or *pol* gene or a region of a retroviral *gag* or *pol* gene for the expression of a retroviral *gag* or *pol* protein in a eukaryotic cell, the expressed retroviral protein being expressed at a level to provide detectable activity of the wild-type

function of the expressed retroviral protein in the eukaryotic cell.

18. The synthetic gene according to claim 17, wherein the retroviral genes have non-preferred codons when referred to the eukaryotic cell, the number of non-preferred codons being such that replacement of all the non-preferred codons by preferred codons with respect to the eukaryotic cell results in a GC nucleotide pair content of 65% or higher, the synthetic gene having a GC nucleotide pair content of between 53 and 63%, more preferably between 55 and 61% and the expressed retroviral protein is expressed at a level to provide detectable enzymatic activity of the expressed retroviral protein in the eukaryotic cell.

19. The synthetic gene according to claim 17 or 18, wherein the expression of the gag or pol proteins is independent of retroviral regulatory proteins.

20. The synthetic gene according to any of the claims 17 to 19, wherein the retroviral protein is a lentiviral gag or pol protein.

21. The synthetic gene according to claim 20, wherein the lentiviral protein is an HIV gag or pol protein.

22. The synthetic gene according to any of claims 18 to 21, wherein the detectable activity of the enzymatic function includes at least promotion or stimulation of the integration of DNA fragments into the host cell DNA, preferably the chromosome of the host cell.

23. The synthetic gene according to any of claims 17 to 22, wherein the retroviral protein is a protease, reverse transcriptase, integrase protein or a polyprotein gag-pol precursor thereof.

24. The synthetic gene according to any of the claims 17 to 23, wherein the eukaryotic cell is a mammalian cell.

25. The synthetic gene according to any of the claims 17 to 24, wherein the expression of

the protein is at a level of at least 200% of that expressed by the wild type gene in the eukaryotic cell.

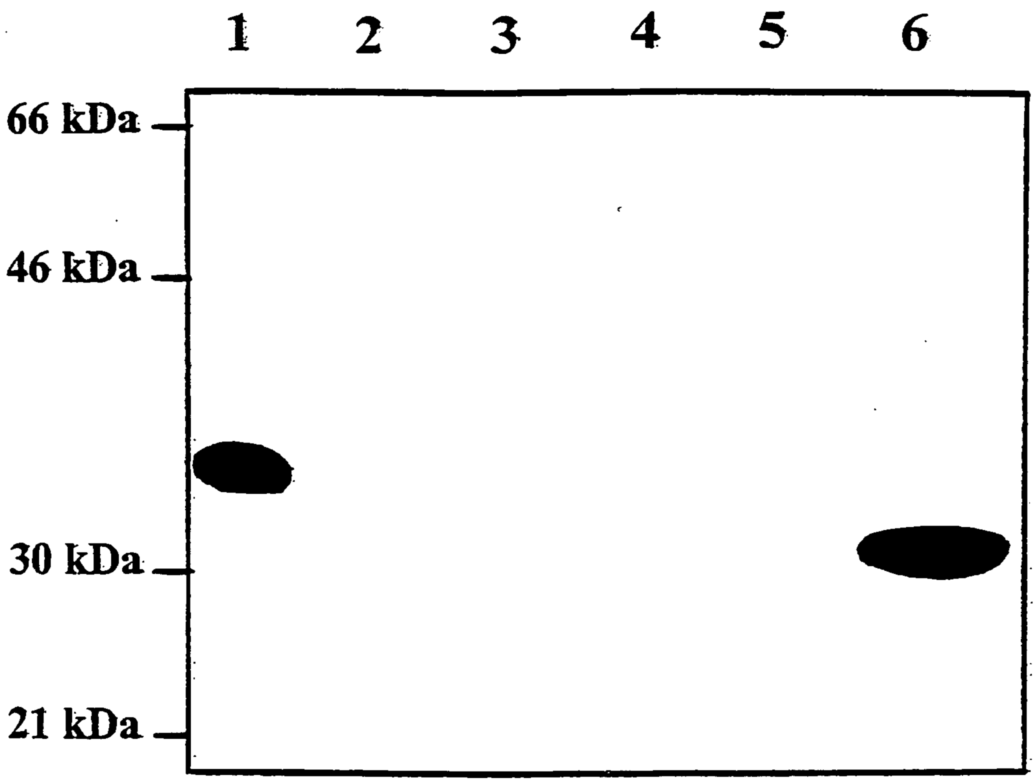
26. The synthetic gene according to any of the claims 17 to 25, comprising the sequence of Fig. 2A or homologs thereof which have a GC content between 53 and 63%, preferably between 55 and 61%.

27. A eukaryotic expression vector comprising the synthetic gene or region of a gene in accordance with any of the claims 17 to 26.

28. The expression vector according to claim 27, further comprising a constitutive or an inducible or a tissue-specific promoter.

29. The expression vector according to claim 27 or 28, comprising a plasmid, a mammalian or an insect virus.

Fig. 1



Lanes: 1. 2.5 ng HT-IN
2. pCEP-4
3. pCEP-IN
4. pCEP-IN-CTE
5. pCEP-IN-RRE + pEF-cRev
6. pCMV-IN^S

Fig. 2A

```
ATCACTAGCA ACCTCAAACA GACACCATGG GATTCCTGGA CGGCATTGAC AAGGCTCAGG AGGAGCACGA GAAGTACCAC TCGAATTGGC GGGCCATGGC   100
                         M G  F L D G  I D K  A Q E  E H E  K Y H  S N W  R  A M A

                                         NheI
CTCCGACTTC AACCTGCCAC CGGTCGTCGC TAAGGAGATC GTTGCTAGCT GCGACAAGTG CCAGCTGAAA GGCGAGGCTA TGCACGGGCA GGTTGATTGC   200
 S D F  N L P  P V V  A K E  I V A  S C D  K C Q  L K G  E A M  H G Q  V D C

TCTCCCGGCA TCTGGCAGCT CGACTGTACT CACCTGGAGG GCAAGGTCAT CCTGGTCGCC GTGCACGTGG CCTCTGGTTA CATCGAGGCT GAGGTCATCC   300
 S P G  I W Q  L D C T  H L E G  K V I  L V A  V H V A  S G Y  I E A  E V I P

PstI
CTGCAGAGAC TGGCCAGGAG ACTGCCTATT TCCTGCTGAA ACTGGCCGGC CGGTGGCCTG TGAAGACAGT GCACACAGAT AACGGCTCCA ACTTCACCTC   400
 A E T  G Q E  T A Y F  L L K  L A G  R W P V  K T V  H T D  N G S N  F T S

                                                 BamHI
CACCACTGTG AAGGCTGCCT GCTGGTGGGC TGGGATCAAG CAGGAGTTCG GGATCCCCTA TAACCCACAG TCTCAGGGCG TGATCGAATC CATGAACAAG   500
 T T V  K A A C  W W A  G I K  Q E F G  I P Y  N P Q  S Q G  V I E  S M N K

GAGCTGAAGA AGATCATCGG CCAGGTTCGG GACCAGGCAG AGCACCTGAA GACTGCAGTG CAGATGGCCG TGTTCATCCA CAACTTCAAG CGAAAGGGGG   600
E L K K  I I G  Q V R  D Q A E  H L K  T A V  Q M A V  F I H  N F K  R K G G

                    NaeI
GCATCGGTGG CTACTCAGCC GGCGAGCGGA TCGTGGACAT CATCGCCACT GACATCCAGA CCAAAGAGCT GCAGAAGCAG ATCACCAAGA TCCAGAACTT   700
 I G G  Y S A  G E R  I V D I  I A T  D I Q T  K E L  Q K Q  I T K I  Q N F

                                                                      NarI
CCGTGTGTAC TACCGGGACT CCCGGGACCC TGTGTGGAAG GGCCCTGCCA AGCTGCTGTG GAAGGGCGAG GGCGCCGTGG TCATTCAGGA CAACTCTGAC   800
 R V Y  Y R D S  R D P  V W K  G P A K  L L W  K G E  G A V V  I Q D  N S D

ATCAAGGTTG TGCCCAGGCG CAAGGCCAAG ATTATCCGGG ACTACGGCAA GCAGATGGCT GGCGACGACT GTGTGGCCTC TCGTCAAGAT GAGGACTAAG   900
I K V V  P R R  K A K  I I R  D Y G K  Q M A  G D D C  V A S  R Q D E D

TCCAACTACT AAACTGGGGG ATATTATGAT                                                                                930
```

Fig. 2B

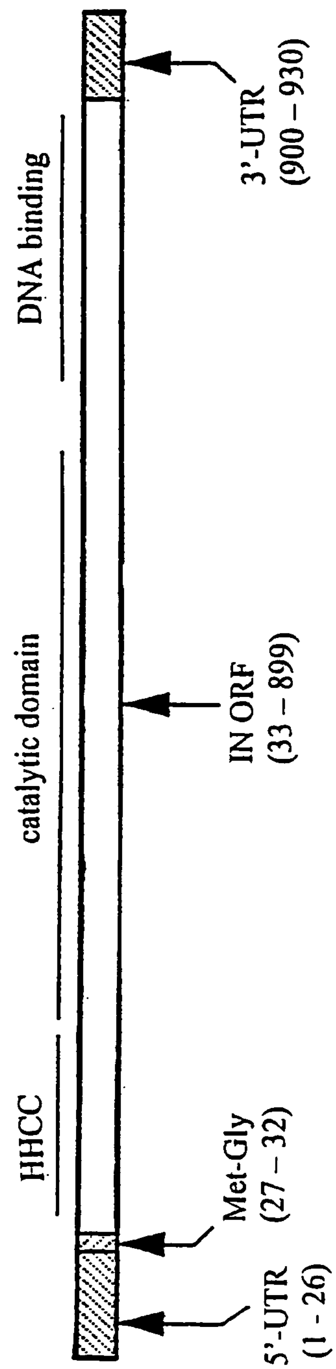
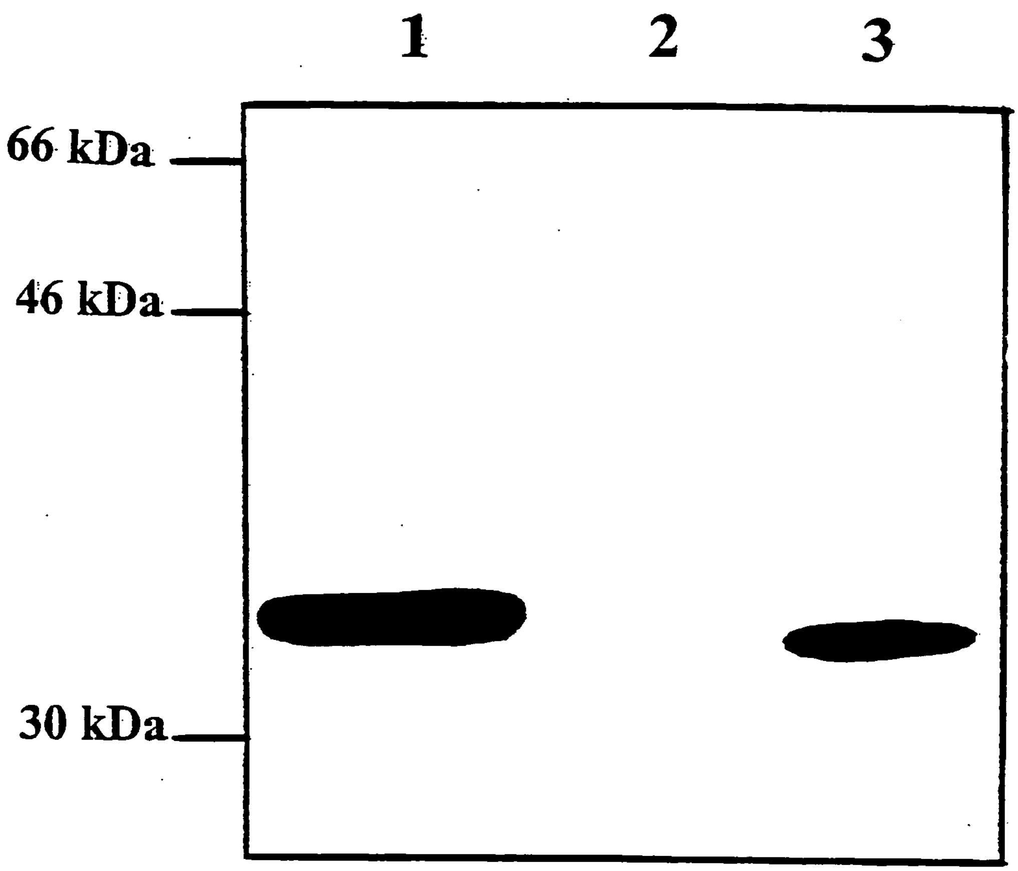

3'-UTR
(900 – 930)
DNA binding
catalytic domain
IN ORF
(33 – 899)
HHCC
Met-Gly
(27 – 32)
5'-UTR
(1 - 26)

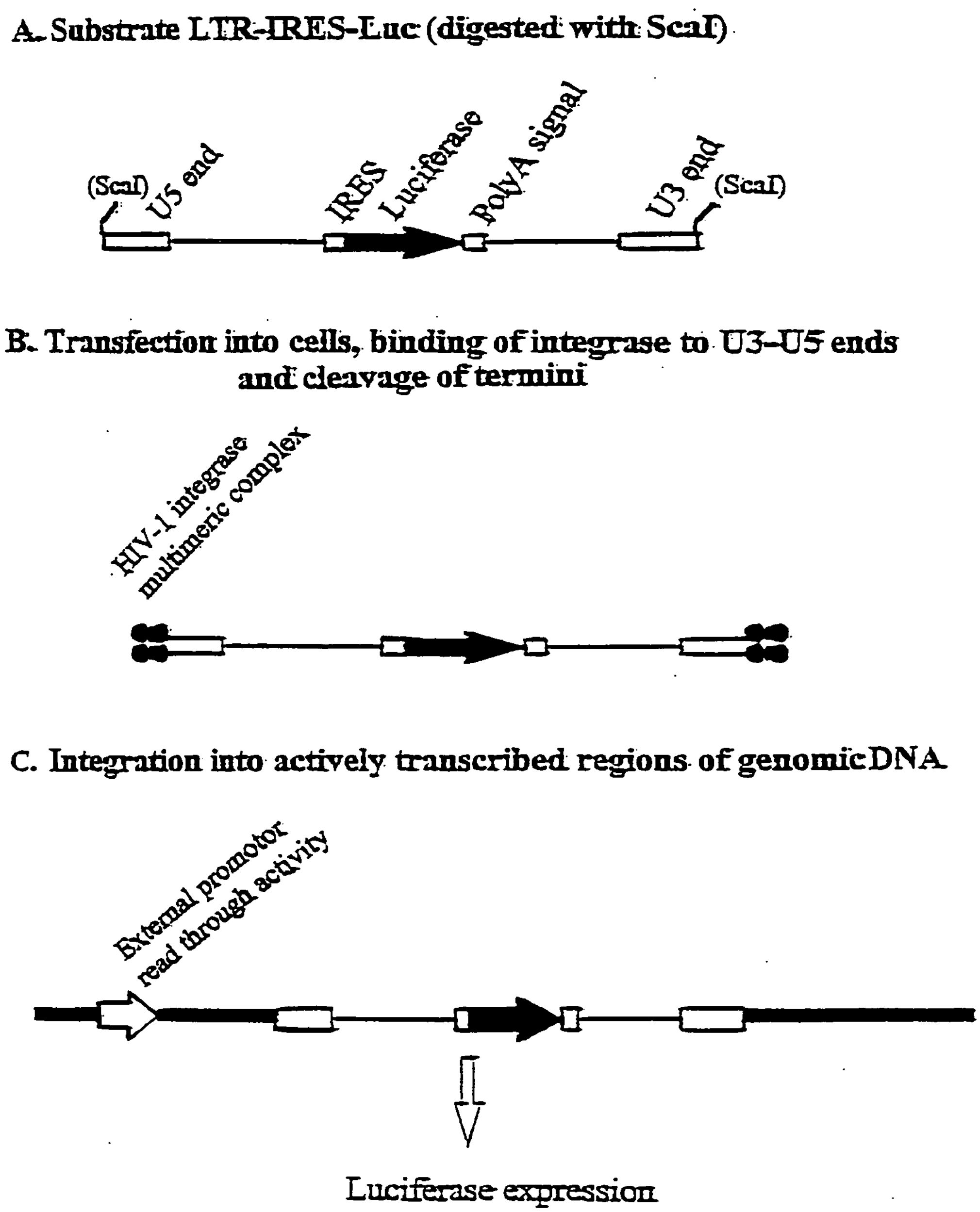
Fig. 3



Lanes: 1. 2.5 ng HT-IN
2. 293T
3. 293T-IN^S

Fig. 4. Principle of DIPR

Detection of integrase activity using a promoterless reporter gene



SEQUENCE LISTING

<110> K .U. Leuven Research & Development
Debyser, Zeger
De Clercq, Erik
Cherepanov, Peter
Pluymers, Wim

<120> A synthetic gene for expression of a retroviral protein
with wild type activity in eukaryotic cells

<130> K1291-PCT

<140>
<141>

<150> EP99201306.0
<151> 1999-04-26

<150> EP00200171.7
<151> 2000-01-18

<160> 2

<170> PatentIn Ver. 2.1

<210> 1
<211> 930
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:synthetic gene
expressing HIV integrase

<220>
<221> CDS
<222> (27)..(899)

<220>
<221> misc_signal
<222> (24)..(30)
<223> Kozak sequence

<400> 1

```
atcactagca acctcaaaca gacacc atg gga ttc ctg gac ggc att gac aag  53
                             Met Gly Phe Leu Asp Gly Ile Asp Lys
                               1               5
```

gct cag gag gag cac gag aag tac cac tcg aat tgg cgg gcc atg gcc 101
Ala Gln Glu Glu His Glu Lys Tyr His Ser Asn Trp Arg Ala Met Ala
10 15 20 25

tcc gac ttc aac ctg cca ccc gtc gtc gct aag gag atc gtt gct agc 149
Ser Asp Phe Asn Leu Pro Pro Val Val Ala Lys Glu Ile Val Ala Ser
30 35 40

tgc gac aag tgc cag ctg aaa ggc gag gct atg cac ggg cag gtt gat 197
Cys Asp Lys Cys Gln Leu Lys Gly Glu Ala Met His Gly Gln Val Asp
45 50 55

tgc tct ccc ggc atc tgg cag ctc gac tgt act cac ctg gag ggc aag 245
Cys Ser Pro Gly Ile Trp Gln Leu Asp Cys Thr His Leu Glu Gly Lys
60 65 70

gtc atc ctg gtc gcc gtg cac gtg gcc tct ggt tac atc gag gct gag 293
Val Ile Leu Val Ala Val His Val Ala Ser Gly Tyr Ile Glu Ala Glu
75 80 85

gtc atc cct gca gag act ggc cag gag act gcc tat ttc ctg ctg aaa 341
Val Ile Pro Ala Glu Thr Gly Gln Glu Thr Ala Tyr Phe Leu Leu Lys
90 95 100 105

ctg gcc ggc cgg tgg cct gtg aag aca gtg cac aca gat aac ggc tcc 389
Leu Ala Gly Arg Trp Pro Val Lys Thr Val His Thr Asp Asn Gly Ser
110 115 120

aac ttc acc tcc acc act gtg aag gct gcc tgc tgg tgg gct ggg atc 437
Asn Phe Thr Ser Thr Thr Val Lys Ala Ala Cys Trp Trp Ala Gly Ile
125 130 135

aag cag gag ttc ggg atc ccc tat aac cca cag tct cag ggc gtg atc 485
Lys Gln Glu Phe Gly Ile Pro Tyr Asn Pro Gln Ser Gln Gly Val Ile
140 145 150

gaa tcc atg aac aag gag ctg aag aag atc atc ggc cag gtt cgg gac 533
Glu Ser Met Asn Lys Glu Leu Lys Lys Ile Ile Gly Gln Val Arg Asp
155 160 165

cag gca gag cac ctg aag act gca gtg cag atg gcc gtg ttc atc cac 581
Gln Ala Glu His Leu Lys Thr Ala Val Gln Met Ala Val Phe Ile His
170 175 180 185

aac ttc aag cga aag ggc ggc atc ggt ggc tac tca gcc ggc gag cgg 629
Asn Phe Lys Arg Lys Gly Gly Ile Gly Gly Tyr Ser Ala Gly Glu Arg
190 195 200

atc gtg gac atc atc gcc act gac atc cag acc aaa gag ctg cag aag 677
Ile Val Asp Ile Ile Ala Thr Asp Ile Gln Thr Lys Glu Leu Gln Lys
205 210 215

cag atc acc aag atc cag aac ttc cgt gtg tac tac cgg gac tcc cgg 725
Gln Ile Thr Lys Ile Gln Asn Phe Arg Val Tyr Tyr Arg Asp Ser Arg
220 225 230

gac cct gtg tgg aag ggc cct gcc aag ctg ctg tgg aag ggc gag ggc 773
Asp Pro Val Trp Lys Gly Pro Ala Lys Leu Leu Trp Lys Gly Glu Gly
235 240 245

gcc gtg gtc att cag gac aac tct gac atc aag gtt gtg ccc agg cgc 821
Ala Val Val Ile Gln Asp Asn Ser Asp Ile Lys Val Val Pro Arg Arg
250 255 260 265

aag gcc aag att atc cgg gac tac ggc aag cag atg gct ggc gac gac 869
Lys Ala Lys Ile Ile Arg Asp Tyr Gly Lys Gln Met Ala Gly Asp Asp
270 275 280

tgt gtg gcc tct cgt caa gat gag gac taa gtccaactac taaactgggg 919
Cys Val Ala Ser Arg Gln Asp Glu Asp
285 290

gatattatga t 930

<210> 2
<211> 290
<212> PRT
<213> Artificial Sequence
<223> Description of Artificial Sequence:synthetic gene
expressing HIV integrase

<400> 2
Met Gly Phe Leu Asp Gly Ile Asp Lys Ala Gln Glu Glu His Glu Lys
1 5 10 15
Tyr His Ser Asn Trp Arg Ala Met Ala Ser Asp Phe Asn Leu Pro Pro
20 25 30
Val Val Ala Lys Glu Ile Val Ala Ser Cys Asp Lys Cys Gln Leu Lys
35 40 45
Gly Glu Ala Met His Gly Gln Val Asp Cys Ser Pro Gly Ile Trp Gln
50 55 60
Leu Asp Cys Thr His Leu Glu Gly Lys Val Ile Leu Val Ala Val His
65 70 75 80
Val Ala Ser Gly Tyr Ile Glu Ala Glu Val Ile Pro Ala Glu Thr Gly
85 90 95

```
Gln Glu Thr Ala Tyr Phe Leu Leu Lys Leu Ala Gly Arg Trp Pro Val
            100                 105                 110
Lys Thr Val His Thr Asp Asn Gly Ser Asn Phe Thr Ser Thr Thr Val
        115                 120                 125
Lys Ala Ala Cys Trp Trp Ala Gly Ile Lys Gln Glu Phe Gly Ile Pro
    130                 135                 140
Tyr Asn Pro Gln Ser Gln Gly Val Ile Glu Ser Met Asn Lys Glu Leu
145                 150                 155                 160
Lys Lys Ile Ile Gly Gln Val Arg Asp Gln Ala Glu His Leu Lys Thr
                165                 170                 175
Ala Val Gln Met Ala Val Phe Ile His Asn Phe Lys Arg Lys Gly Gly
            180                 185                 190
Ile Gly Gly Tyr Ser Ala Gly Glu Arg Ile Val Asp Ile Ile Ala Thr
        195                 200                 205
Asp Ile Gln Thr Lys Glu Leu Gln Lys Gln Ile Thr Lys Ile Gln Asn
    210                 215                 220
Phe Arg Val Tyr Tyr Arg Asp Ser Arg Asp Pro Val Trp Lys Gly Pro
225                 230                 235                 240
Ala Lys Leu Leu Trp Lys Gly Glu Gly Ala Val Val Ile Gln Asp Asn
                245                 250                 255
Ser Asp Ile Lys Val Val Pro Arg Arg Lys Ala Lys Ile Ile Arg Asp
            260                 265                 270
Tyr Gly Lys Gln Met Ala Gly Asp Asp Cys Val Ala Ser Arg Gln Asp
        275                 280                 285
Glu Asp
    290
```